DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

A. JAMES CLARK SCHOOL OF ENGINEERING

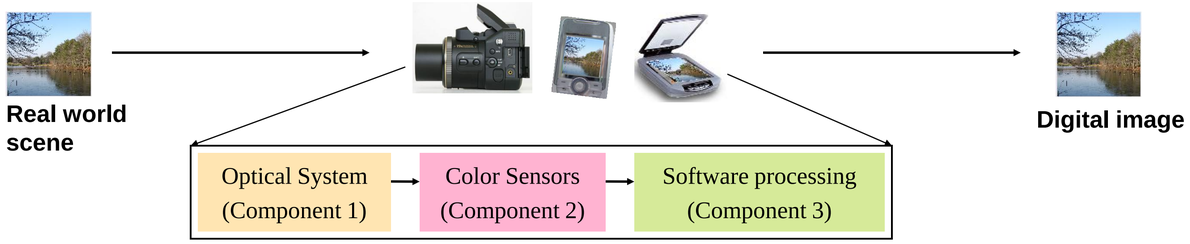
Digital Image Forensics

Ashwin Swaminathan, Hongmei Gou, Wei-Hong Chuang, Christine McKay, Min Wu, and K.J. Ray Liu

Images contain intrinsic traces…

How is a digital image created? What type of device captured the image? What are inside the capture device? Has the image been manipulated after capture? How?

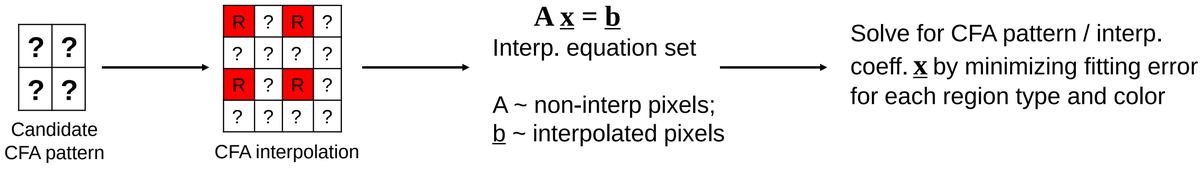
We developed methodologies to answer various forensic questions, such as image source classification, device brand / model identification and tampering detection.



Each component in a digital device modifies the input via a certain algorithm and leaves intrinsic traces in the final output that can be extracted to make forensic inferences.

Identifying source devices

Algorithms / parameters are used in digital devices are estimated to identify the devices; e.g., digital cameras use different Color Filter Arrays (CFAs) for scene sampling and color interpolation. Exact CFA pattern and interpolation coefficients can be estimated.



Algorithm / Parameter Estimation

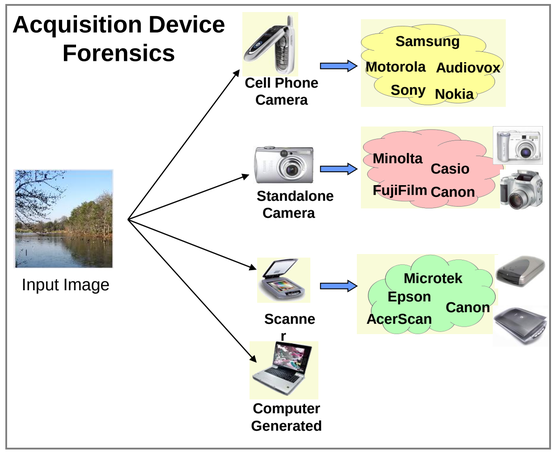
Canon EOS

Fujifilm S3000

0	-0.005	0	0.008	0	0.001	0
0.018	0	-0.079	0	-0.050	0	0.006
0	-0.021	0	0.315	0	-0.023	0
0.003	0	0.325	1.000	0.301	0	0.023
0	-0.023	0	0.301	0	-0.015	0
-0.006	0	-0.061	0	-0.051	0	0.010
0	0.002	0	0.018	0	-0.001	0

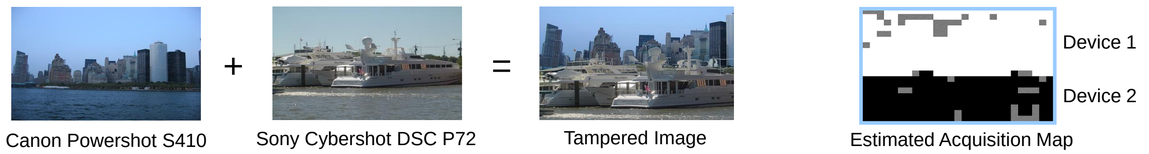
Smooth region coefficients from Canon A75

Acquisition Device Forensics



Detecting tampering and manipulations

- Camera inconsistency utilized to detect cut-and-paste tampering

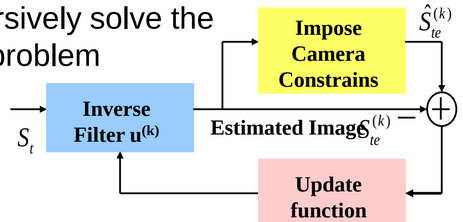


Canon Powershot S410 | Sony Cybershot DSC P72 | Tampered Image | Estimated Acquisition Map

- A universal method for detecting “post-capture” manipulations
  1. Approximate the manipulation with an LSI filter and find filter coefficients **u** by minimizing estimation error

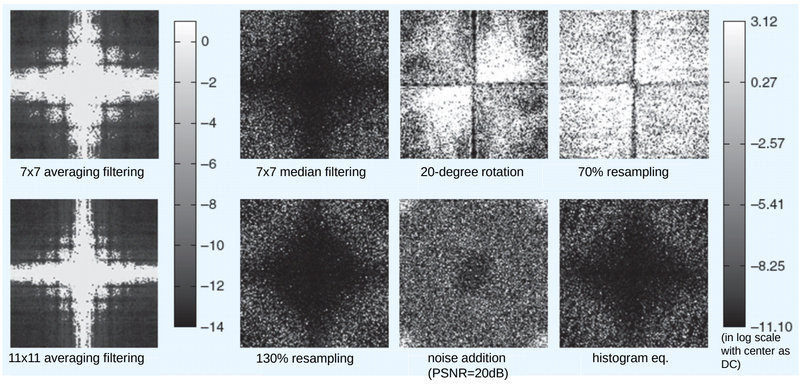
$$J = \sum_{x,y,c} \left(\hat{S}_{te}(x,y,c) - S_{te}(x,y,c) \right)^2 + \eta \sum_{c=1}^{3} \left(\sum_{m,n} u(m,n,c) - 1 \right)^2$$

  2. Recursively solve the min. problem

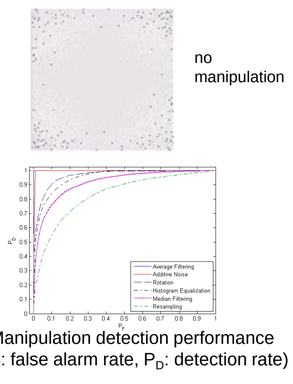


  3. Estimated filter = identity ⟶ image not manipulated

Typical estimated “frequency responses”



≠



Manipulation detection performance (P_F: false alarm rate, P_D: detection rate)

Distinguishing manipulations

- Empirical frequency responses (EFRs) associated with images undergoing the same manipulation are often clustered, exploited to classify different manipulations, including non-linear, spatially-varying ones.